

Dynamic Alteration of the Loss Function in Convolutional Neural Networks for Medical Image Segmentation

Vinicius Ferreira de Souza¹, Alexei M. C. Machado^{1,2}

¹Department of Computer Science
Pontifical Catholic University of Minas Gerais (PUC Minas)
30.535-901 – Belo Horizonte – MG – Brasil

²Department of Anatomy and Imaging – Universidade Federal de Minas Gerais
Av Alfredo Balena 190 – 30130-100 – Belo Horizonte – MG – Brasil

qsilva@sga.pucminas.br, alexei@sga.pucminas.br

Abstract. *This work presents a method for training convolutional neural networks (CNNs) based on the dynamic alteration of the loss function that guides model optimization. First, a comparative analysis of six different loss functions applied to CNNs for medical image segmentation is conducted. Through experiments with the colon cancer dataset from the Medical Segmentation Decathlon, distribution-based functions (Focal Loss, TopK Loss, and Binary Cross-Entropy) and region-based functions (Dice Loss, IoU Loss, and Tversky Loss) are evaluated. The results demonstrate that distribution-based functions consistently outperform region-based functions. Furthermore, the dynamic alteration of the functions achieved an increase in accuracy of up to 12% when compared to composite approaches. Qualitative analysis revealed that this approach not only improves quantitative metrics but also enhances the visual quality of segmentation. These findings provide valuable insights for optimizing CNN performance in medical image segmentation and contribute to the development of more accurate solutions in AI-assisted diagnosis.*

1. Introduction

The advancement of Computer Vision techniques has significantly transformed healthcare, enabling the automation of important tasks ranging from early anomaly detection to artificial intelligence-assisted diagnostics. In this context, Convolutional Neural Networks (CNNs) have stood out for some time as powerful tools in medical image segmentation, such as X-rays, Magnetic Resonance Imaging, and Ultrasound, enabling the precise identification of anatomical structures and pathologies.

A fundamental aspect, yet frequently underestimated, of the success of these models is the appropriate choice of the loss function, which directly influences training efficiency and the accuracy of the final model. The loss function guides the learning process, determining how a neural network adjusts its weights to minimize the error between predictions and ground truth values. In the medical context, where images frequently exhibit significant class imbalances, the choice of loss function becomes even more important.

The present work proposes a CNN training method based on the dynamic alteration of the loss function that guides the optimization of models applied to the problem of medical image segmentation. Through systematic experiments on the colon cancer

dataset from the Medical Segmentation Decathlon, three distribution-based functions (Binary Cross-Entropy, Focal Loss, and TopK Loss) and three region-based functions (Dice Loss, IoU Loss, and Tversky Loss) are evaluated regarding their performance and robustness in binary segmentation tasks.

The main contributions of this work include: a method that dynamically alters the loss function of the gradient backpropagation algorithm; a comparative analysis of distinct loss functions categorized into distribution-based and region-based functions; the investigation of strategies involving the dynamic alteration of the loss function as an alternative to traditional composite functions, demonstrating its superiority; and a detailed qualitative and quantitative evaluation that reveals both numerical and visual improvements in segmentation quality. An analysis of related works suggests that the idea of dynamically altered loss functions has not yet been reported in the literature, which highlights the contribution of the present work.

2. Related Works

The definition of loss functions that guide the optimization process of deep network training can be considered one of the keys to the success of these architectures applied to image classification and segmentation across various domains, having inspired numerous initiatives. An example can be found in the work of [Lin et al. 2020], which introduced the Focal Loss in the context of RetinaNet for object detection, aiming to learn more complex examples and reduce the impact of false negatives on easy examples. The function uses a modulating factor that allows the model to concentrate its learning on the most challenging pixels, which is an important characteristic in medical images, where structures of interest are frequently small and subtle. The TopK Loss, proposed by [Wu et al. 2016], selects only the $k\%$ pixels with the lowest predicted probability for the target class, proving effective in scenarios with noisy annotations.

[Milletari et al. 2016] adapted the Dice Coefficient as a differentiable loss function, the Dice Loss, which is effective for the segmentation of small objects. The IoU Loss, explored by [Rahman and Wang 2016], offers an alternative with slightly different characteristics in the penalization of false positives and negatives. [Salehi et al. 2017] introduced the Tversky Loss, generalizing the Tversky index to independently control the weight of false positives and false negatives, a flexibility particularly valuable in medical applications.

The categorization of loss functions was systematized by [Ma et al. 2021], who proposed a taxonomy with four categories: distribution-based, region-based, boundary-based, and composite. The authors demonstrate that composite functions (particularly Dice + TopK Loss) offer greater robustness than single functions, providing empirical evidence that motivates the investigation of combination strategies in the present work. The papers by [Wang et al. 2020] and [Tian et al. 2022] present complementary reviews on loss functions in machine learning.

Although recent advances in medical image segmentation have explored Transformer-based architectures, the choice of loss function remains a fundamental aspect of training regardless of the underlying architecture, which motivates the present investigation.

A loss scheduling technique for addressing class imbalance problems in semantic

image segmentation was presented by [Taubert et al. 2020]. The method is a composite approach in which the weights change gradually during training. Two scheduling strategies were proposed: Naïve (linear transition) and Alternating (linear ramp followed by sinusoidal decay). Experiments on three datasets (PASCAL VOC 2012, SBD, and Cityscapes) demonstrated that loss scheduling could consistently direct the model to local minima that emphasize underrepresented classes, improving class-averaged recall compared to traditional static loss functions while maintaining adequate calibration properties. However, the work presented limitations in the systematic definition of the weights and scheduling hyperparameters, opting for simplified heuristics instead of specific optimization strategies for each dataset and loss function combination.

The strategy presented in this paper differs from the previous literature in the sense that it proposes an abrupt change in the loss function, from a region-based to a distribution-based approach, instead of linearly combining multiple loss functions for training.

3. Loss Functions

All loss functions presented in this study were selected based on the works of [Ma et al. 2021] and [Xu et al. 2023]. The choice of these six functions and their combinations is justified by their prevalence in studies found in the image segmentation literature. Thus, three distribution-based functions (Focal Loss, TopK Loss, and Binary Cross-Entropy) and three region-based functions (Dice Loss, IoU Loss, and Tversky Loss) were chosen for the experiments with composite functions from different categories, analyzing their impacts on the final segmentation metric results. The motivation for choosing these functions was to investigate whether the complementary characteristics of each category could be effectively combined through dynamic loss function switching.

In the remainder of this text, I denotes an image in R^2 or R^3 domains, and S and G are the segmentation result and its ground truth, respectively. In this notation, s_i and g_i correspond to the predicted segmentation and the ground truth of voxel i , N is the number of voxels in image I , and C is the number of classes.

3.1. Distribution-Based Functions

Distribution-based loss functions aim to minimize the dissimilarity between two distributions. The primary function in this category is Cross-Entropy, from which all others are derived.

Binary Cross-Entropy: adapts Information Theory to measure differences between probability distributions in binary classification tasks [Xu et al. 2023], computing the error pixel by pixel through the negative log-likelihood function and being especially effective when there is class imbalance. A more easily implementable version is the Binary Cross-Entropy with Logit Loss, which applies the sigmoid function after the original function, defined as

$$L_{BCE} = -\frac{1}{N} \sum_{i=1}^N \gamma_i \log(s_i) + (1 - \gamma_i) \log(1 - g_i), \quad (1)$$

where $\gamma \in \{0, 1\}$ is the binary class of the i -th pixel.

Focal Loss: represents an adaptation of conventional cross-entropy that concentrates learning on challenging samples, reducing the weight assigned to easily classified examples, while simultaneously addressing the class imbalance problem between object and background [Lin et al. 2020]. It is defined as:

$$L_{Focal} = -\frac{1}{N} \sum_c^C \sum_{i=1}^N (1 - s_i^c)^\gamma g_i^c \log s_i^c, \quad (2)$$

where $\gamma > 0$ reduces the relative loss for well-classified examples. The original paper indicates the value of $\gamma = 2$ as the most effective.

TopK Loss: an extension of the cross-entropy loss function designed specifically to handle hard examples during training [Wu et al. 2016]. The TopK Loss drives the model to learn solely based on the pixels that are most difficult to classify at each iteration. The $k\%$ of pixels with the lowest predicted probability for the target class are selected, and only their loss is considered. It is defined by:

$$L_{TopK} = -\frac{1}{N} \sum_c^C \sum_{i \in K} g_i^c \log s_i^c, \quad (3)$$

where K is the set of the $k\%$ worst pixels.

3.2. Region-Based Functions

Region-based functions aim to minimize the disparity or maximize the overlap between regions of the true labels g and the predicted segmentation s . The most popular function in this category is the Dice Loss.

Dice Loss: adapts the Dice similarity metric as a loss function to maximize the intersection between predictions and ground truth annotations [Drozdal et al. 2016], providing better performance in the segmentation of small objects and cases with unequal class distribution. It is defined as:

$$L_{Dice} = 1 - \frac{2 \sum_{c=1}^C \sum_{i=1}^N g_i^c s_i^c}{\sum_{c=1}^C \sum_{i=1}^N g_i^c + \sum_{c=1}^C \sum_{i=1}^N s_i^c}. \quad (4)$$

IoU (Jaccard) Loss: adapts the Jaccard index as a differentiable loss function to directly optimize the Intersection over Union metric, measuring the ratio between the intersection and the union of the segmented masks, demonstrating superiority over pixel-based losses for semantic segmentation tasks [Rahman and Wang 2016]. It is defined as:

$$L_{IoU} = 1 - \frac{2 \sum_{c=1}^C \sum_{i=1}^N g_i^c s_i^c}{\sum_{c=1}^C \sum_{i=1}^N (g_i^c + s_i^c - g_i^c s_i^c)}. \quad (5)$$

Tversky Loss: adapts the Tversky index in the form of a generalized loss function to address data imbalance in medical segmentation, introducing the hyperparameters α and β to independently control the weight of false positives and false negatives, enabling a better trade-off between precision and recall [Salehi et al. 2017]. It is defined as:

$$L_{Tversky} = 1 - T(\alpha, \beta) \quad (6)$$

$$= 1 - \frac{\sum_c^C \sum_{i=1}^N g_i^c s_i^c}{\sum_c^C \sum_{i=1}^N g_i^c s_i^c + \alpha \sum_c^C \sum_{i=1}^N (1 - g_i^c) s_i^c + \beta \sum_c^C \sum_{i=1}^N g_i^c (1 - s_i^c)}, \quad (7)$$

where α and β are hyperparameters that compensate for the weight of false negatives and false positives.

3.3. Composite Functions

Composite functions are the combination of two or more functions simultaneously, aiming to aggregate their impacts. These functions were used as a comparison in the second experimental hypothesis to evaluate whether the approach of switching functions during training could outperform the use of traditional composite functions. In this work, composite functions consisting of 2 terms were analyzed, in the form $L = L_A + L_B$:

- **Dice Focal Loss:** is the sum of the Dice and Focal functions in order to alleviate class imbalance and force the model to better learn from poorly segmented pixels.
- **IoU Binary Cross-Entropy Loss:** combines the optimization of the Intersection over Union metric with the robustness of binary cross-entropy.
- **Tversky Binary Cross-Entropy Loss:** combines the fine-grained control of false positives and negatives from Tversky Loss with the robustness of binary cross-entropy.
- **IoU TopK Loss:** combines the optimization of the IoU overlap metric with a focus on the most challenging pixels from TopK Loss.

The functions were taken from the PyTorch and segmentation-models-pytorch libraries. The Tversky Loss used $\alpha = 0.3$, and $\beta = 0.7$; the Focal Loss was implemented with $\gamma = 2$, and the TopK Loss with $k = 0.6$. All other functions used the default parameters available in the PyTorch and segmentation-models-pytorch libraries.

4. Materials and Methods

The methodology of this work was structured to allow for a systematic and comparative evaluation of different loss functions in the context of medical image segmentation, their simultaneous combinations, and dynamic alteration.

4.1. Model Architecture

The segmentation model used to implement the proposed dynamic alteration of the loss function was based on the U-Net architecture with an EfficientNet-B0 backbone and is shown in Fig. 1. The U-Net architecture, proposed by [Ronneberger et al. 2015], is characterized by its skip connections that preserve high-resolution spatial information during the decoding process. This architecture became an important reference for medical segmentation due to its ability to efficiently handle the scarcity of annotated data, which is common in medical applications. The EfficientNet model was proposed by

[Tan and Le 2019] and generated a family of architectures that simultaneously optimize network depth, width, and resolution through compound scaling. This choice was motivated by the computational efficiency of EfficientNet, which offers an optimized balance between parameters and learning capacity, and by the inherent adaptability of the U-Net skip connections for segmenting complex medical structures.

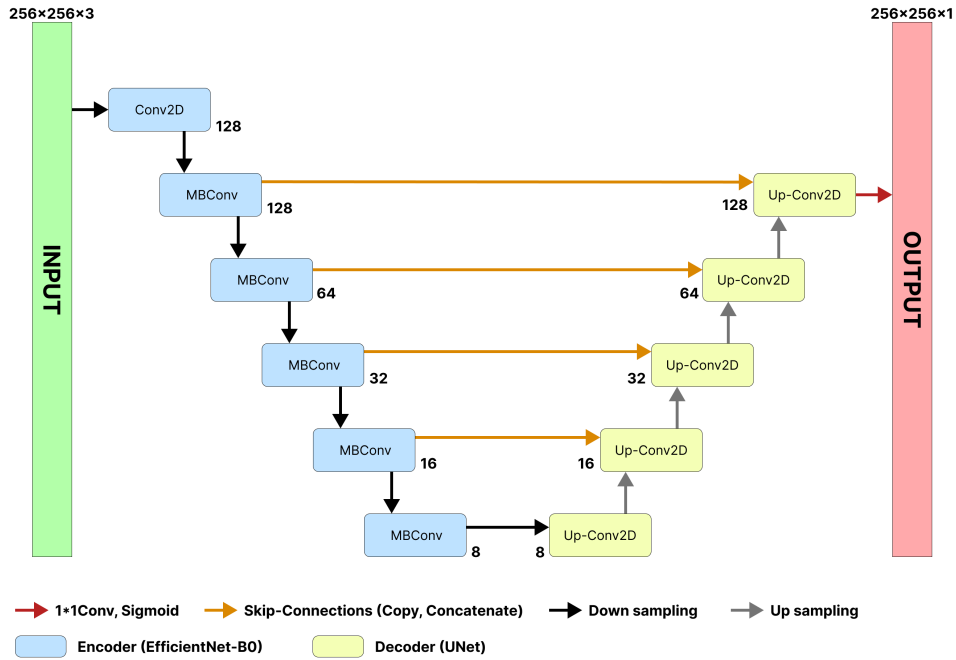


Figure 1. Architecture of the segmentation network used to implement the proposed dynamic loss alteration.

The model was configured with 3 input channels (RGB) and 1 output channel with a sigmoid activation function. The hyperparameters used in the U-Net model training were: a batch size of 32; learning rate of $1e-4$ with adaptive reduction; and the Adam optimizer.

4.2. Dataset

The dataset used in this article was the tenth task of the Medical Segmentation Decathlon [Antonelli et al. 2022], specifically focused on colon cancer segmentation. The dataset contains 13,486 images with a resolution rescaled to 256×256 voxels per plane and corresponding binary masks. The choice of this dataset was based on three main criteria: (1) usability and ease of processing; (2) substantial quantity of images that allows robust training; and (3) binary nature of segmentation, suitable for demonstrating the results of different loss functions in a controlled and well-defined scenario. The dataset presents a ratio of 9.5:1 (1,285 images containing the target class and 12,201 containing only background) at the image level, characterizing a moderately imbalanced scenario that allows for adequate evaluation of the performance of different proposed approaches. The dataset was partitioned into 10,788 images for training and 2,698 for testing. Images from the same patient were kept either in the training or the testing set to avoid data leakage.

4.3. Evaluation Metrics

The main metrics used to evaluate the performance of the image segmentation models were the Dice coefficient and the IoU metric, which quantify the similarity between model predictions and true labels (ground truth), providing indicators of segmentation quality.

4.3.1. Dice Coefficient

The Dice Coefficient, also known as the Sørensen-Dice Index or F1-Score, is a metric widely used to measure similarity between two datasets, and it being especially popular in image segmentation tasks and machine learning model evaluation. It is defined as

$$Dice(G, S) = 2|G \cap S|/(|G| + |S|). \quad (8)$$

4.3.2. Intersection over Union

The IoU is one of the most important and widely used metrics for evaluating image segmentation quality and object detection in computer vision and machine learning. It is defined as

$$IoU(G, S) = |G \cap S|/|G \cup S|. \quad (9)$$

5. Experiments and Results

The experiments were conducted on a local i7-13700F 16/24, 5.2 GHz processor, with 32 GB of RAM and an RTX 4070 12 GB GPU. To evaluate the model performance with different functions, IoU and Dice metrics were used for each test performed. The method was implemented using the PyTorch framework due to its flexibility and efficiency in developing deep learning models.

The first set of tests used single loss functions for network training. The values presented in Table 1 correspond to the average Dice and IoU metrics obtained in the experiments for 50 training epochs. The number of epochs was defined beforehand based on the duration of training (an average of approximately 40 minutes) and was shown to be sufficient for convergence.

The second set of tests investigated the use of composite functions. Table 2 presents the average results obtained by composite functions during 100 epochs, i.e., functions from different approaches applied simultaneously throughout training.

The third set of tests investigated the dynamic alteration of region-based loss functions to distribution-based functions, aiming to evaluate whether the impact on performance metrics would be cumulative or if learning from the first function would be replaced by the second. Unlike the loss scheduling work of [Taubert et al. 2020], which transitions from dice loss to cross-entropy aiming to improve performance on imbalanced classes, our proposal substitutes a region loss for a distribution loss with the objective of increasing model accuracy regardless of imbalance. Experiments were conducted using different combinations of loss functions, where the first function served as a base for the initial training for 50 epochs, followed by the application of the second function for an additional 50 epochs, totaling 100 epochs.

Table 1. Average±standard deviation of the Dice and IoU metrics for the experiments with single loss functions.

Loss Function	Dice	IoU
Focal loss	0,7538 ± 0.0062	0,6342 ± 0.0060
TopK loss	0,7465 ± 0.0142	0,6265 ± 0.0154
Binary Cross Entropy loss	0,7462 ± 0.0111	0,6324 ± 0.0140
Dice loss	0,6976 ± 0.0132	0,5792 ± 0.0134
IoU (Jaccard) loss	0,6759 ± 0.0151	0,5582 ± 0.0164
Tversky loss	0,6740 ± 0.0290	0,5469 ± 0.0329

Table 2. Average±standard deviation of the Dice and IoU metrics for the experiments with composite loss functions.

Loss Function	Dice	IoU
Tversky + Binary Cross Entropy	0,7153 ± 0.0088	0,5985 ± 0.0103
IoU + Binary Cross Entropy	0,7060 ± 0.0030	0,5912 ± 0.0027
IoU + TopK	0,7034 ± 0.0026	0,5897 ± 0.0023
Dice + Focal	0,6983 ± 0.0018	0,5828 ± 0.0019

The methodological strategy was based on the hypothesis that different categories of loss functions have complementary characteristics that could be exploited through the dynamic alteration of different functions, offering an alternative to traditional linear combinations. Thus, three region-based functions (IoU, Dice, and Tversky) were used as a base for initial training, followed by the application of three distribution-based functions (Focal, TopK, and Binary Cross-Entropy) as refinement after 50 epochs, totaling 9 combinations, which allowed for the evaluation of the interactions between these two distinct optimization strategies in image segmentation. Table 3 presents the results obtained through the dynamic alteration of loss functions.

Table 3. Average±standard deviation of the Dice and IoU metrics for the experiments with Dynamic Alteration.

Dynamic Alteration	Dice	IoU
Dice loss followed by Focal loss	0,7642 ± 0.0054	0,6507 ± 0.0067
Dice loss followed by TopK loss	0,7734 ± 0.0068	0,6608 ± 0.0081
Dice loss followed by Binary Cross Entropy	0,7758 ± 0.0083	0,6574 ± 0.0083
IoU loss followed by Focal loss	0,7769 ± 0.0035	0,6558 ± 0.0036
IoU loss followed by TopK loss	0,7851 ± 0.0078	0,6689 ± 0.0077
IoU loss followed by Binary Cross Entropy	0,7882 ± 0.0052	0,6750 ± 0.0054
Tversky loss followed by Focal loss	0,7583 ± 0.0056	0,6426 ± 0.0063
Tversky loss followed by TopK loss	0,7842 ± 0.0089	0,6667 ± 0.0084
Tversky loss followed by Binary CrossEntropy	0,7866 ± 0.0122	0,6754 ± 0.0129

6. Discussion

A qualitative analysis of the loss function behavior during training and validation, as well as the predicted segmentation masks obtained from the third group of experiments, re-

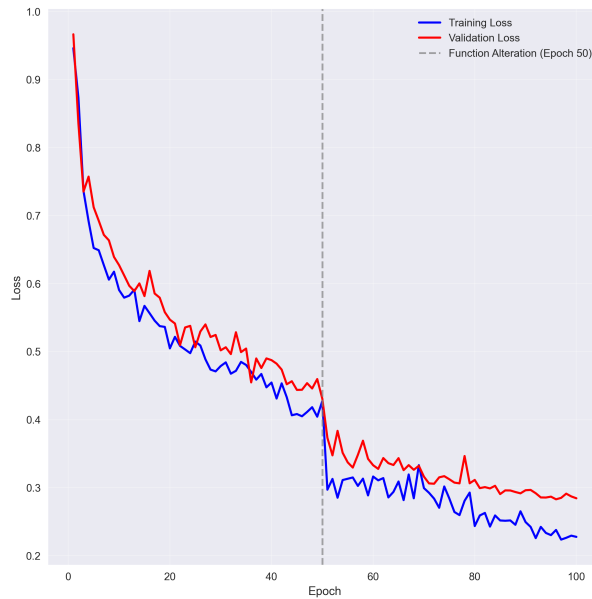


Figure 2. Illustration of the training/validation loss behavior when the loss function is dynamically altered from IoU Loss to Dice Loss at epoch 50.

veals a significant benefit achieved by the dynamic alteration of the loss function. The example in Fig. 2 uses a transition between two region-based functions (IoU to Dice) to illustrate the switching mechanism. This choice was necessary because the substantial difference in magnitude between region-based losses (typically ranging from 0.3 to 1.0) and distribution-based losses (operating at values below 0.01) prevents meaningful visualization of a cross-category transition in a single plot. The sudden drop pattern shown in Fig. 2 is consistently observed across all tested combinations, including the region-to-distribution transitions reported in Table 3. This behavior suggests that altering the loss function during training allows for combining the advantages of different approaches, resulting in more precise and robust segmentation. Fig. 3 shows an example in which the model trained with dynamic alteration, starting with a region-based loss function and subsequently refined with TopK Loss, improved prediction quality and corrected specific errors observed when using single or composite functions. The improvement in segmentation is clear, with both a reduction in prediction errors and an enhancement in the precision of tumor delineation.

Regarding clinical relevance, the Dice score of 0.78 obtained by the dynamic alteration approach can be contextualized against the Medical Segmentation Decathlon challenge, where colon cancer was one of the hardest tasks, with a median Dice of 0.16 across participants and 0.56 for the top-ranking method [Antonelli et al. 2022]. In medical image segmentation, Dice scores above 0.7 are generally considered acceptable for computer-aided screening, while scores above 0.9 are typically required for clinical deployment. The obtained results thus suggest potential applicability in a screening context, where automated segmentation serves as a preliminary delineation subject to expert review.

In order to assess the statistical significance of the improvement obtained by dynamic alteration, a 5-fold cross-validation set of additional experiments was performed. The same partition proportion of 80% of the data for training and 20% for testing was

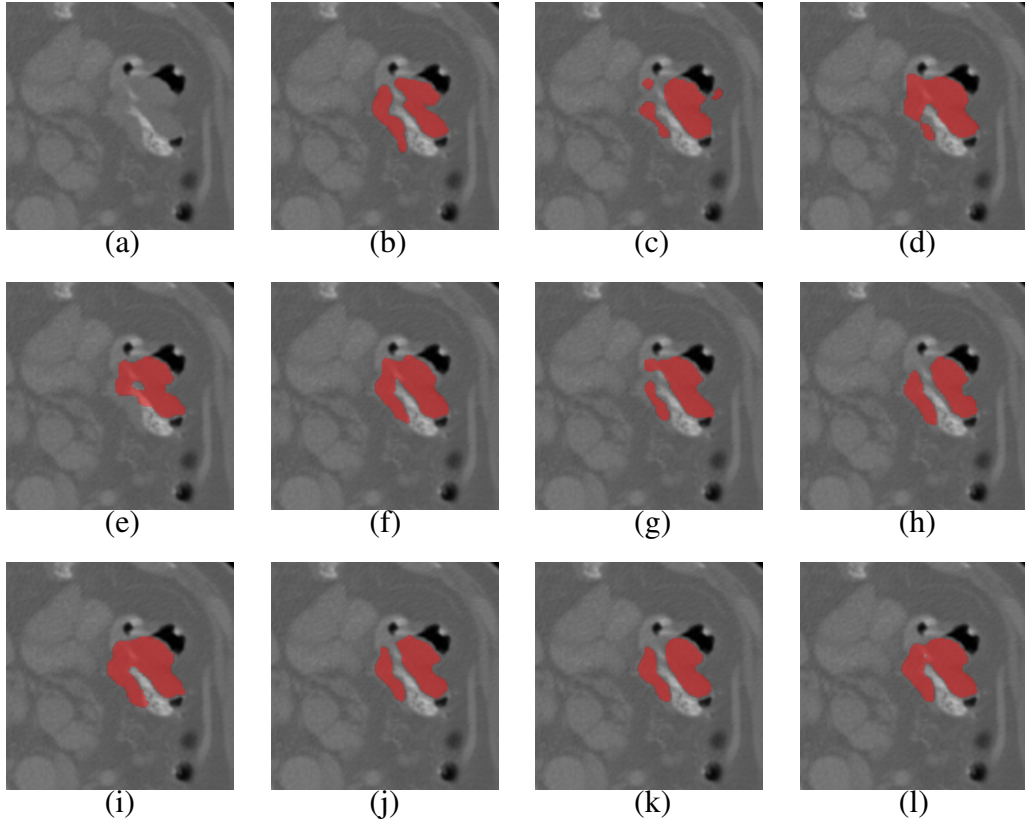


Figure 3. Comparison of the results obtained for the segmentation of slice 102 of image 008 from Medical Decathlon. (a) 150×150 pixel region of interest of the original image, starting at $x,y=(200,100)$, showing the tumor and (b) Ground-truth segmentation mask. The remaining images show the segmentation obtained using: IoU loss (c); Dice loss (d) Tversky loss (e); Topk loss (f); IoU + TopK (g); Dice + Topk (h); Tversky + Topk (i); IoU followed by TopK (j); Dice followed by Topk (k); and Tversky followed by Topk (l).

applied to the composite functions and to the dynamic alteration. The objective was to determine whether results from the linear combination of two loss functions in a single training stage would be surpassed by the results obtained through dynamic alteration, where the second loss function is applied after training the model with the first function. For the composite functions, the model was trained in a single step for 100 epochs. In the case of the dynamic alteration, the model was trained with the first loss function for 50 epochs, followed by the second function for an additional 50 epochs. Two-tailed Student's t tests were used to assess the significance of the differences obtained in both approaches.

The results from Table 4 consistently demonstrate that the dynamic alteration of the loss function surpasses the use of composite functions in all tested combinations. Differences in Dice metrics ranged from 0.09 to 0.15, representing improvements of 13 to 24% in performance. For the IoU metric, the differences ranged from 0.08 to 0.13, a gain between 14 and 24%. All p-values were smaller than 0.003.

The superiority of dynamic function alteration over composite functions can be explained by the following factors:

1. *Gradual Specialization*: Altering from one function to another during training

Table 4. Mean(standard deviation) and t -statistics(p-value) for the 5-fold cross-validation experiments comparing composite function (CF) and dynamic alteration (DA) based on Dice and IoU metrics.

Loss functions	Dice Performance Metric			IoU Performance Metric		
	CF	DA	t (p-value)	CF	DA	t (p-value)
Tversky and BCE	0.66 (0.03)	0.78 (0.02)	7.44 (0.0001)	0.54 (0.03)	0.67 (0.02)	8.06 (0.0001)
IoU and BCE	0.65 (0.03)	0.77 (0.03)	6.32 (0.0002)	0.56 (0.03)	0.67 (0.03)	5.80 (0.0004)
IoU and TopK	0.63 (0.02)	0.78 (0.02)	11.86 (0.0001)	0.54 (0.02)	0.67 (0.02)	10.23 (0.0001)
Dice and Focal	0.67 (0.03)	0.76 (0.02)	5.58 (0.0005)	0.57 (0.03)	0.65 (0.03)	4.22 (0.0029)

allows the network to first learn general characteristics with one function, then to specialize with the second function.

2. *Convergence Stability*: It avoids conflicts between potentially antagonistic objectives of the two functions during training when one loss is added to the other, as in the composite approach.
3. *Solution Space Exploration*: It enables more efficient exploration of parameter space in two distinct stages.

7. Conclusion

This study presented a strategy for training CNNs using two consecutive loss functions instead of single or composite ones. Experiments conducted on the colon cancer segmentation task revealed that distribution-based functions showed superior performance compared to region-based functions for the analyzed colon cancer dataset. The dynamic function alteration approach proved to be an effective strategy for combining the advantages of different categories of loss functions. Experiments evidenced that starting training with region-based functions and subsequently changing to distribution-based functions resulted in consistent improvements in quantitative metrics. The IoU followed by TopK and the Tversky followed by BCE achieved the best overall performance in tested cases (Dice: 0.78; IoU: 0.67).

Qualitative analysis of the predicted masks revealed that, in addition to numerical improvements, this approach provided visually more accurate segmentation, with a significant reduction in prediction errors and better definition of the edges of the segmented objects. Particularly notable was the correction of specific segmentation problems observed while using single loss functions, such as the incorrect segmentation of objects into separate regions.

Based on the obtained results, some limitations should be acknowledged: the evaluation was conducted on a single dataset, the switching point at epoch 50 was not systematically optimized, and the comparison between single functions (50 epochs) and dynamic alteration (100 epochs) does not fully isolate the effect of extended training. However, the comparison between composite and dynamic approaches confirms the superiority of the proposed method. Future work includes validation on additional medical datasets, experimentation with alternative switching points, and the development of adaptive strategies that automatically determine when to transition between loss functions based on convergence metrics.

Acknowledgments — Alexei M. C. Machado acknowledges the financial support

from CNPq and Fundação de Amparo à Pesquisa de Minas Gerais (FAPEMIG) grants APQ-02753-24 and APQ-06556-24.

References

- Antonelli, M., Reinke, A., Bakas, S., Farahani, K., Kopp-Schneider, A., Landman, B. A., Litjens, G., Menze, B., Ronneberger, O., Heller, N., et al. (2022). The medical segmentation decathlon. *Nature Communications*, 13(1):4128.
- Drozdal, M., Vorontsov, E., Chartrand, G., Kadoury, S., and Pal, C. (2016). The importance of skip connections in biomedical image segmentation. In *Deep Learning and Data Labeling for Medical Applications*, pages 179–87. Springer.
- Lin, T.-Y., Goyal, P., Girshick, R., He, K., and Dollár, P. (2020). Focal loss for dense object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(2):318–27.
- Ma, J., Chen, J., Ng, M., Huang, R., Li, Y., Li, C., Yang, X., and Martel, A. L. (2021). Loss odyssey in medical image segmentation. *Medical Image Analysis*, 71:102035.
- Milletari, F., Navab, N., and Ahmadi, S.-A. (2016). V-net: Fully convolutional neural networks for volumetric medical image segmentation.
- Rahman, M. A. and Wang, Y. (2016). Optimizing intersection-over-union in deep neural networks for image segmentation. In *Advances in Visual Computing*, pages 234–44. Springer.
- Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention*, pages 234–41.
- Salehi, S. S. M., Erdogmus, D., and Gholipour, A. (2017). Tversky loss function for image segmentation using 3d fully convolutional deep networks. In *Machine Learning in Medical Imaging*, pages 379–87. Springer.
- Tan, M. and Le, Q. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. In *International Conference on Machine Learning*, volume 97, pages 6105–14.
- Taubert, O., Götz, M., Schug, A., and Streit, A. (2020). Loss scheduling for class-imbalanced image segmentation problems. In *IEEE International Conference on Machine Learning and Applications*, pages 426–31.
- Tian, Y., Su, D., Lauria, S., and Liu, X. (2022). Recent advances on loss functions in deep learning for computer vision. *Neurocomputing*, 497:129–58.
- Wang, Q., Ma, Y., Zhao, K., and Tian, Y. (2020). A comprehensive survey of loss functions in machine learning. *Annals of Data Science*, pages 1–26.
- Wu, Z., Shen, C., and van den Hengel, A. (2016). Bridging category-level and instance-level semantic image segmentation.
- Xu, H., He, H., Zhang, Y., Ma, L., and Li, J. (2023). A comparative study of loss functions for road segmentation in remotely sensed road datasets. *International Journal of Applied Earth Observation and Geoinformation*, 116:103159.